Understanding face familiarity

Robin S. S. Kramer[1,2], Andrew W. Young[1], A. Mike Burton[1]

[1] Department of Psychology, University of York, UK

[2] School of Psychology, University of Lincoln, UK

Correspondence to:

A. Mike Burton,

Department of Psychology,

University of York,

York,

YO10 5DD, United Kingdom.

mike.burton@york.ac.uk

Abstract

It has been known for many years that identifying familiar faces is much easier than identifying unfamiliar faces, and that this familiar face advantage persists across a range of tasks. However, attempts to understand face familiarity have mostly used a binary contrast between 'familiar' and 'unfamiliar' faces, with no attempt to incorporate the vast range of familiarity we all experience. From family members to casual acquaintances and from personal to media exposure, familiarity is a more complex categorisation than is usually acknowledged. Here we model levels of familiarity using a generic statistical analysis (PCA combined with LDA) computed over some four thousand naturally occurring images that include a large variation in the numbers of images for each known person. Using a strong test of performance with entirely novel, untrained everyday images, we show that such a model can simulate widely documented effects of familiarity in face recognition and face matching, and offers a natural account of the internal feature advantage for familiar faces. Furthermore, as with human viewers, the benefits of familiarity seem to accrue from being able to extract consistent information across different photos of the same face. We argue that face familiarity is best understood as reflecting increasingly robust statistical descriptions of idiosyncratic within-person variability. Understanding how faces become familiar appears to rely on *both* bottom-up statistical image descriptions (modelled here with PCA), and top-down processes that cohere superficially different images of the same person (modelled here with LDA).

Keywords: Face recognition; familiarity; face matching; face learning

1. Introduction

The concept of familiarity is central to our understanding of face recognition. It has been known for many years that perception of familiar and unfamiliar faces differs in a number of ways (for reviews see Johnston & Edmonds, 2009; Young & Burton, 2017), and this point is emphasised in theoretical models (Bruce & Young, 1986; Burton, Bruce & Hancock, 1999). For example, in studies of recognition memory, familiar faces are recognised faster and more accurately than unfamiliar faces (Ellis, Shepherd & Davies, 1979; Klatzky & Forrest, 1984; Yarmey, 1971). This difference is not in any straightforward sense purely a memory effect, because in more recent studies of perceptual face *matching*, participants are again more accurate with familiar (compared to unfamiliar) faces, when judging whether two images depict the same person (e.g. Bruce et al., 1999, 2001; Burton, Wilson, Cowan & Bruce, 1999; Megreya & Burton, 2006, 2008).

Despite these differences, our working definition of familiarity has been unsophisticated and our understanding of what happens when a face becomes increasingly familiar has been limited at best. Almost all studies compare unfamiliar, never previously seen, faces to highly familiar people, often well-known celebrities. However, our daily experience tells us that familiarity is not simply a dichotomy. We all know many people with varying levels of familiarity, from members of our family encountered every day over long periods, to casual acquaintances perhaps seen occasionally on our route to work, or serving us in an infrequently-visited café. In this paper, we aim to capture familiarity in all its diversity. We present a model of face recognition which incorporates a large range of familiarity, and explore the consequences of increasing familiarity.

One key effect of familiarity is that it leads to generalisable representations for recognition. Early memory studies consistently showed that superficial image changes in pose, expression or lighting were detrimental to memory for unfamiliar faces, but had very little effect on familiar face memory (e.g. Bruce, 1982; Hill & Bruce, 1996;

O'Toole, Edelman & Bülthoff, 1998; Patterson & Baddeley, 1977). This has led to the idea that unfamiliar face processing is highly image-bound (Hancock, Bruce & Burton, 2000; Megreya & Burton, 2006). In consequence, recognition declines as a function of differences between study and test photos (Beveridge et al., 2011; Estudillo & Bindemann, 2014), since representations of unfamiliar faces are tied to the specific images that were encountered. This image-dependence for unfamiliar faces seems to hold even after extensive training involving repeated exposure to a small number of different views of the same face (Liu, Bhuiyan, Ward & Siu, 2009; Longmore, Liu & Young, 2008). In such circumstances, particular training examples themselves become well-recognised, but show little generalisation to novel examples of the learned faces.

In marked contrast to unfamiliar face recognition, recognition of highly familiar faces is very robust. We can tolerate severe image degradation (Burton et al, 1999; Bruce et al, 2001) and considerable image distortion (Hole, George, Eaves & Rasek, 2002) with very little effect on our ability to recognise the people we know. Why might this be? One proposal that lies at the heart of the approach we develop here is that our exposure to familiar faces has itself been highly diverse, including the very wide variability in the appearance of any particular individual that arises under everyday conditions (Jenkins, White, van Montfort & Burton, 2001; Burton 2013; Jenkins & Burton, 2011). To illustrate this point, consider Figure 1, comprising five photos of the actor Hugh Jackman. These pictures vary due to characteristics of the person (e.g. age, hairstyle, weight), the pose and facial expression, the image capture conditions (e.g. lighting, viewpoint) and the capture device (e.g. perspective settings, exposure levels). The images are therefore superficially very different in a way that is typical of everyday, ambient images (Burton, Jenkins & Schweinberger, 2011). However, despite this diversity, a viewer familiar with the actor can recognise Hugh Jackman easily in all the photos. Our proposal in earlier work has been that this is because we have already encountered his face in a wide range of conditions, allowing us to have built up a representation of him which *includes* information about the ways in which his face can vary.

Figure 1. Unconstrained ambient images of the same person. Depicted variation is due to changes in pose, lighting, expression, age, camera settings, and so on. Image attributions from left to right: Eva Rinaldi (Own work) [CC BY-SA 2.0], Grant Brummett (Own work) [CC BY-SA 3.0], Gage Skidmore (Own work) [CC BY-SA 3.0], Eva Rinaldi (Own work) [CC BY-SA 2.0], Eva Rinaldi (Own work) [CC BY-SA 2.0].

The nature of face representations has, of course, been a long-standing concern. In particular, many researchers have asked how it might be possible to build a representation that can be accessed when presented with *any* recognisable instance of a particular face (Bruce & Young, 1986; Eger, Schweinberger, Dolan & Henderson, 2005). Most conceptions, until recently, have emphasised what might potentially be common to all images of a person. For example, the most widely used idea involves the second-order configuration of distances between facial features (Carey & Diamond, 1977), though this is now known to run into both empirical and conceptual difficulties (Burton, Schweinberger, Jenkins & Kaufmann, 2015; Maurer, Le Grand & Mondloch, 2002). Alternatively, it has been pointed out that there might be common texture patterns across the face that can be captured through image averaging (Burton, Jenkins, Hancock & White, 2005). Such approaches imply, at least implicitly, that familiarisation results in higher fidelity representations which can become sufficiently refined to be recruited when recognising a novel image of a known person. By focusing on what might be common to all views of the same face, research in this tradition thus often treats within-person variability – the extent to which the same face can look different – as noise. Typical experimental approaches in consequence tend to use highly controlled stimuli in which images of different people are taken under very similar conditions (lighting, pose, expression, camera).

The approach used here represents a break from this tradition. We have recently followed an important insight of Bruce (1994) and suggested that, rather than being irrelevant noise, within-person variability can actually *assist* in finding information that is diagnostic of individual identity (Burton, Kramer, Ritchie & Jenkins, 2016). This is because statistical analysis of multiple images of the same person shows that within-person variability is, to some extent, idiosyncratic. So, the ways in which one face varies are different from the ways in which another varies. Under this proposal, it is important to sample widely over different, naturally occurring images of someone in order to become familiar with that person - because part of familiarisation is learning that person's unique variability.

This proposal that variability is central to creating effective representations of face identities is gaining experimental support. For example, participants learn a face more effectively when exposed to greater variation in the images they see (Menon, White & Kemp, 2015a; Murphy, Ipser, Gaigg & Cook, 2015; Ritchie & Burton, 2017). So, while traditional approaches to face learning emphasise image-independent factors such as *duration* of exposure (Read, Vokey & Hammersley, 1990; Reynolds & Pezdek, 1992), this may not be so critical as the image-dependent *type* of exposure, and especially the *range* of exposure. Likewise, if people have idiosyncratic facial variability, then we would expect any training on a particular face to have rather limited generalisability to other faces. Once again, this is borne out by experiments studying training in face recognition. Facial learning can be enhanced by various training regimes, but the benefits accrue only to those faces encountered, and do not generalise to others (Dowsett Sandford & Burton, 2016; Hussain, Sekuler & Bennett, 2009).

Renewed interest in face learning, as described above, highlights the fact that we need a better understanding of familiarity. Studies manipulating levels of familiarity do so, almost exclusively, through a binary categorisation of faces as 'familiar' or 'unfamiliar', and tests of learning tend to dichotomise responses as 'seen' or 'unseen'. An exception is a series of experiments by Clutterbuck and Johnston (2002, 2004, 2005)

who show that pairwise matching – i.e. the ability to match two different images of a face – varies relatively smoothly with levels of familiarity. Nevertheless, for the most part, familiarity is treated in the research literature as a discontinuous variable with only two states.

In this paper, we take the important step of examining familiarity as a multi-valued function. We present a development of a previously implemented computational model (Kramer, Young, Day & Burton, 2017a) using minimal assumptions and standard image analysis techniques involving a combination of Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA). Whilst this approach has already been shown to simulate the specific property of image invariant familiar face recognition (Kramer et al., 2017a), these methods potentially offer a generic approach to exploiting statistical regularities in the images. Here, we show that the same approach can be used to simulate a range of key properties of face recognition across different levels of familiarity.

Building on earlier research, Kramer et al. (2017a) demonstrated that combining PCA with LDA is effective at capturing the human-like property of good recognition of novel views of familiar faces when the training involves a substantial number of images of each face. To achieve this result, however, Kramer et al. (2017a) used an implementation that involved training their model on a fixed number of instances of each face. In this sense, their approach was based on a specific combination of circumstances in which some faces were uniformly familiar (trained across the same number of images) and other faces were completely unfamiliar (untrained). This of course approximates the binary way in which familiarity has often been conceptualised in the research literature. Here we take the further critical steps necessary to arrive at a more general understanding of familiarity, building a model in which some identities are represented by only a single photo, whereas others are represented by varying numbers of different photos, creating a parallel with the different degrees of familiarity encountered in everyday life. Unlike many traditional models based on a single standardised view of each face, we train the model to recognise people based on prior exposure to widely varying images of each face and evaluate its performance with a strong test involving entirely novel, untrained and

highly varied ambient images. The basics of the approach we use are exactly the same as those used by Kramer et al. (2017a); only the composition of the training or test image sets is changed.

Computer simulation offers the considerable advantage of forcing the theorist to make every aspect of a model fully explicit, but it also carries the attendant risk of crafting a model that 'works' only under the specific set of circumstances for which it was created. The best way to mitigate this risk is to demonstrate that the model can encompass phenomena that extend well beyond those from which it was derived (Young & Burton, 1999). To show that our extension of Kramer et al.'s (2017a) approach does indeed offer general insights into face familiarity we used it to simulate a range of key findings from the face recognition literature. We demonstrate not only that our model benefits from being trained across more exemplars but also that this increases resistance to the effects of image degradation and can account for the finding that increasing familiarity particularly enhances recognition based on the face's internal rather than external features (Ellis, Shepherd & Davies, 1979; Young, Hay, McWeeny, Flude, & Ellis, 1985). As a further demonstration of the model's applicability, we show that it can encompass findings from widely used face matching tasks.

Having demonstrated the model's wide applicability, we finish by investigating in more detail what lies behind these findings and their implications for understanding the nature of face familiarity. We show how LDA reshapes the perceptual space created by PCA, and that the benefits of familiarity are largely but not entirely specific to each familiar face. Moreover, we examine the importance of supervised learning to this process. Our approach involves a combination of an unsupervised 'bottom-up' analysis (PCA of the image training set) with supervised 'top-down' learning (via LDA) of the characteristics of a set of trained identities. Supervised learning approximates what happens in everyday life in that we will usually know who someone is during a social encounter. By comparing the resulting PCA and PCA+LDA spaces we investigate how far face recognition can be based on the unsupervised image statistics of the perceptual input alone (via PCA) and to what extent it benefits from a combination of top-down with

bottom-up influences (PCA+LDA). These observations have broad implications for understanding the nature of perceptual expertise with faces.

2. The model

We begin by implementing a basic model to demonstrate that increasing familiarity with a face (as indexed by the number of different photos of the face on which the model is trained) differentially enhances the recognition of new (untrained) images of that face. Having established this parallel with behavioural demonstrations of image invariance for recognition of familiar faces, we turn to investigating whether the same model can account for resistance to image degradation in recognising familiar faces, for the internal feature advantage for familiar face recognition, and for performance in face matching tasks. Finally, we explore in more detail what happens in the model as a face becomes increasingly familiar.

2.1. Image sets

In order to model real-world exposure to faces, we collected ambient, everyday images. These were similar in nature to the 'Labeled Faces in the Wild' database (Huang, Ramesh, Berg, & Learned-Miller, 2007), which attempts to incorporate natural variability across numerous dimensions, including pose, lighting, expression, age, and camera conditions. We used images in which no part of the face was obscured (by clothing, glasses, hands, etc.). To facilitate the placement of landmark fiducial points on each image, we also limited our image poses to within approximately ±30° from full face. Beyond these limits some fiducials would be obscured; for example when one edge of the face is no longer seen as the view moves toward profile. Apart from these minimal technical requirements, the face images were entirely unconstrained.

Based on these criteria, we collected a large set of 4,154 colour images using Google Image search. These images included 335 different identities, where the majority

were White but other ethnicities featured, and approximately half were women. Many were Hollywood actors, although people from other professions (athletes, politicians, etc.) were also represented. The 'level of familiarity' was represented by varying the number of images of different faces in the set, ranging from a single image (for 161 identities) up to 159 images for the most 'highly familiar' individual. For the remaining identities, the number of images per face varied widely: $M = 22.16$ images, $SD = 26.20$. In all cases, we simply took the first n images (where n was the number of images required) returned by Google Image search that met the pose criteria given above. In this way we sought to ensure that as far as possible the images would reflect the variability that might be encountered for each face.

Images were cropped to include only the head, rescaled to 190 pixels wide x 285 pixels high, and represented in RGB colour space using a lossless image format (bitmap).

2.2. General procedure

We used LDA to train our model to group different images of the same person together. This technique fits exemplars (here faces) to a space in which intra-class differences are minimised, while inter-class differences are maximised, i.e. faces of the same person are clustered together. This is a technique which has been used in many previous models of face recognition (e.g. Belhumeur, Hespanha & Kriegman, 1997; Jing, Wong & Zhang, 2006; Kramer et al., 2017a) and is sometimes referred to as the Fisherface approach because the discriminant function used is due to R.A. Fisher (1936). When classifying images, it is common to have fewer sample vectors (images) than features (pixels). In such cases, LDA cannot be carried out without first reducing the number of feature dimensions. This can be done in a number of ways, including morphological analysis of faces to create a reduced-dimensional description (e.g., Chen et al., 2000). A more popular approach is first to subject the faces to Principal Components Analysis (PCA), resulting in a low-dimensional description of 'eigenfaces' representing the variability in the image set (e.g., Bekios-Calfa, Buenaposada & Bamela, 2011). In our studies, we adopted this PCA-based approach to dimension reduction, as follows.

All images were shape-standardised by morphing them to a template derived from the average shape of the entire set (Burton, Miller, Bruce, Hancock & Henderson, 2001; Craw, 1995). This standardisation was based on the alignment of 82 fiducial points for each image (e.g., corners of eyes, corners of mouth etc.; for technical details, see Burton et al., 2016, and for downloadable face processing software, see Kramer, Jenkins, & Burton, 2017b). Assignment of these fiducial points was carried out using a standard semi-automatic process requiring just five manually-entered landmarks (see Kramer et al., 2017b, for details). PCA was then computed on these normalised images. In order to reduce the number of dimensions describing the resulting space without significant loss of variability, we retained the highest 335 components only. This corresponds to the number of identities and is therefore the minimum number of PCs required for the subsequent LDA. These principal components explained 95.6% of the variance in the image RGB information. The images' projections on these principal components were then entered into an LDA, where each class represented an identity. The result is a reshaped space comprising 334 dimensions (the number of identities minus 1). Again, to reduce the number of dimensions describing the PCA+LDA space without significant loss in performance, we retained the first 143 components, which accounted for 95.0% of the 'discriminability' from the overall LDA space.

Our face identity training involved applying this PCA+LDA procedure to a large set of training images in order to produce a space that could best distinguish the 335 identities. The actual size of the training set was subject to minor variations when essential to address specific questions, as noted below. For example, a small proportion of the 4,154 available images was left untrained when these were needed to serve as novel test items for familiar face recognition (e.g. in Section 3.1, below).

Although our analysis was based primarily on shape-normalised images, we do not wish to imply that face shapes are unimportant in recognition or familiarization. Normalised faces still carry considerable information about the shape of the original – for example shape-from-shading cues are retained in gradients of intensity within the image.

While some previous work has attempted to separate shape from texture (e.g. Andrews et al, 2016; Itz et al., 2017), we remain neutral here about their relative influence, and about their relative contribution to the normalized representations we employ. In order to establish whether the normalization process removes useful information, however, we also ran the PCA+LDA procedure described above on the shape vectors comprising raw fiducial points of the original images and on the shape vectors calculated in terms of differences from the average fiducial position. We describe these simulations below.

3. Simulations

3.1. Familiarity improves recognition

A basic requirement for a model of familiarity is that recognition is more accurate for more familiar faces. Novel, previously unseen (i.e. untrained) images of well-known faces should be better recognised than novel images of less well-known people. To examine this, we first sampled one image of each to-be-trained person to act as untrained 'test' images. This was possible for all the identities represented by at least two images (i.e. 174 identities). We then ran the PCA+LDA procedure, as described above, without these test photos (i.e. with 3,980 training images). Next, we projected each untrained test image into the resulting space, and computed its distance from representations of the known faces. This procedure was repeated for 100 iterations, each time randomly selecting the image of each identity to be used as the test image. Model accuracy was then calculated by averaging responses across all iterations, producing a proportion of iterations in which the model was correct.

There are two common ways to judge successful recognition in this type of model. One is to measure the distance between the test item and all other images, counting the 'nearest neighbour' image as the model's 'decision', with this being correct or incorrect if the nearest neighbour is a photo of the same (correct) or a different (incorrect) identity. Alternatively, we can calculate a centroid for each known identity – i.e. the mean position

of all its exemplars in the PCA+LDA space – and represent the model's decision as the nearest centroid.

We examined both of these metrics because they can be seen as approximating instance-based approaches to face recognition in which recognition will be in terms of the most similar variant encountered before (the nearest neighbour measure), such as Hay (2000), or the more abstractive type of model in which recognition is based on a representation that can generalise across many instances (the nearest centroid measure), such as Bruce and Young's (1986) concept of 'face recognition units' or Burton et al.'s (2005) image averaging approach. Distinguishing between instance-based and abstractive accounts has not been easy in terms of behavioural data (e.g. Ellis, Young, Flude & Hay, 1987; Hay, 2000; Young & Bruce, 2011), and we also found that the nearest neighbour and nearest centroid measures generated highly comparable patterns. We therefore chose to focus on the nearest centroid measure here for two reasons. First, the nearest centroid approach approximates what remains the dominant theoretical perspective (e.g. Bruce & Young, 1986). Second, creating a single centroid for each known face avoids the danger of inflating the recognition rate at low levels of performance through the possibility of random 'hits' resulting from the presence of multiple instances for the more familiar faces. While we have chosen a centroid approach to decisions in the model, we do not wish to exclude other possibilities, and so in the simulations below we also report summary statistics for the nearest neighbour measure when it can be used, to illustrate that the pattern is always the same.

Figure 2 shows the relationship between familiarity (i.e. number of training images) and the proportion of 'correct' identifications of untrained test images, based on the nearest centroid measure. Each data point in Figure 2 represents the average proportion of correct decisions across test images for a specific identity. As can be seen, some faces are easier to recognise than others, even at low levels of familiarity (i.e. with few training images); this corresponds to the well-known phenomenon of facial distinctiveness (Valentine, 1991, Valentine & Bruce, 1986). More importantly there is a clear, highly significant association between familiarity and recognition rate, such that novel photos of

increasingly familiar people are correctly recognised more often than novel photos of less familiar people, $r_s(172) = .74$, $p < .0001$. For the nearest neighbour measure, rank correlation gave a closely comparable value of $r_s(172) = .72$, $p < .0001$.

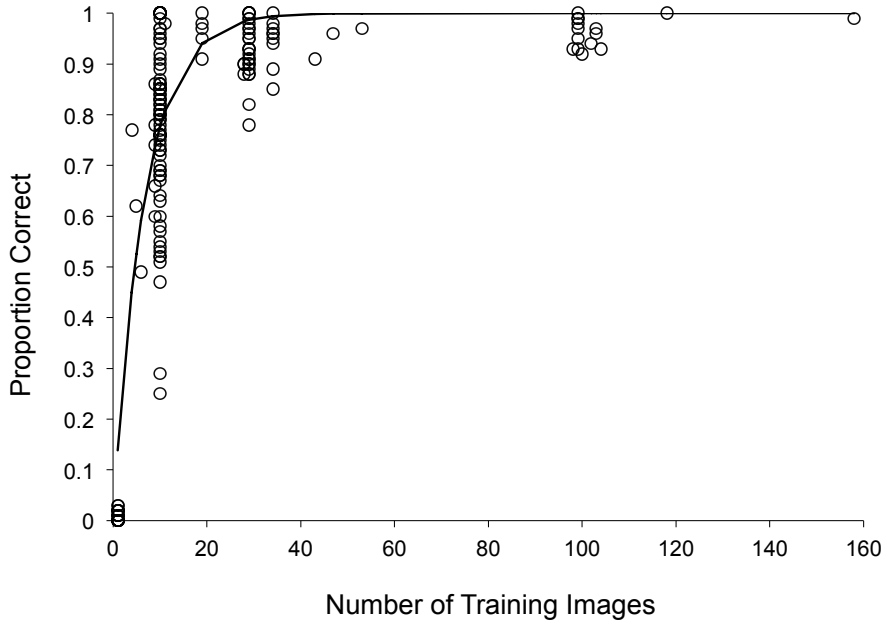

Figure 2. The proportion of correct recognitions of untrained novel face images using the 'nearest centroid' measure increases for more familiar faces (where familiarity is represented by the number of training images). Each point represents the average across test images for a specific identity (but note that several of these data points are overlapping). The fitted curve represents an exponential function. Although some faces are easier to recognise than others, even at low levels of familiarity, there is a substantial correlation between familiarity and recognition rate ($r_s = 0.74$), such that novel photos of increasingly familiar people are correctly recognised more often than novel photos of less familiar people.

A pitfall that needs to be avoided in statistical learning studies is that of overfitting,

in which the model finds essentially spurious random patterns in the data. To guard against overfitting, the data presented in Figure 2 used a strong test of recognition based on correct classification of novel (untrained) images of the target faces. As an additional precaution, however, we ran simulations in which PCA was carried out as usual but the image identities were randomly scrambled at the LDA stage. This gives data of the same order, but with no top-down structure, i.e. it attempts to use LDA to cluster together random sets of images of different people. If such a model were nevertheless able to learn an effective categorization, this would provide evidence against the utility of our approach and instead imply that overfitting remains possible despite the precaution of using novel test images. In fact, and reassuringly, this procedure caused performance to collapse completely, resulting in mean recognition rates of 0.01 for both centroid and nearest neighbour measures. This collapse in performance shows that random patterns in the data are of little use in classifying the identities of these highly variable face images.

As noted above, we also evaluated whether familiarity can improve recognition based only on the 2D shape information given by the raw positions of the fiducials in each image, or by their differences from the average locations. A combined PCA+LDA of the locations of the 82 fiducial points in the unstandardised images showed that performance was very poor (mean recognition rates of 0.03 for both centroid and nearest neighbour measures), and it remained poor when we applied the same technique to shape vectors calculated in terms of differences from the average fiducial position (mean recognition rates still 0.03 for both centroid and nearest neighbour measures). These poor levels of performance do not of course show that 2D shape information is irrelevant, but they do show that the information about fiducial locations in each image removed by our normalization procedure is of no value over and above the normalized information itself.

3.2. Familiarity confers resistance to image degradation

As well as being recognisable across many different views in normal conditions, familiar faces can also often be recognised from degraded images such as those created

by low resolution video surveillance cameras (Burton et al, 1999; Bruce et al, 2001). To operationalise image degradation in our model, we constructed images where a proportion of the pixels were replaced by the average RGB pixel values of the entire set, as illustrated in Figure 3. This manipulation results in images with the same dimensions as the whole set, which can therefore be used to test recognition in exactly the same way. However, those pixels that were replaced by the average values become completely uninformative for face identity.

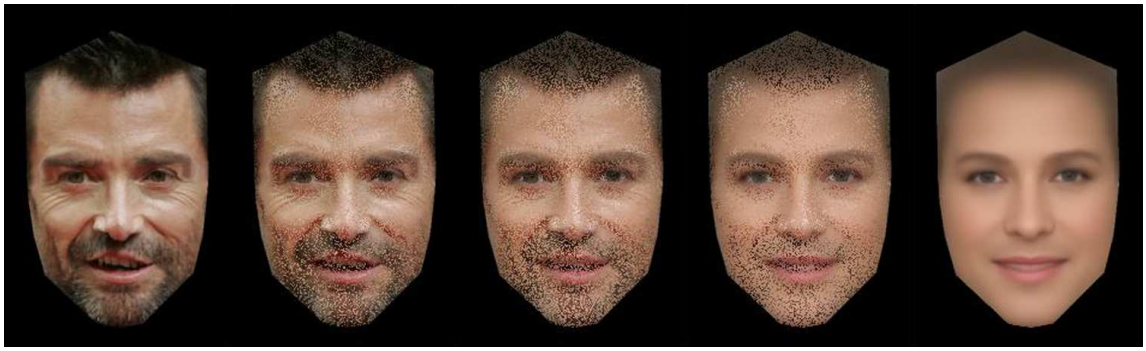

Figure 3. An illustration of the image degradation manipulation. A shape-standardised full-face image is shown at the far left, and the average of all images from the training set at the far right. Intermediate images have 25%, 50%, and 75% of their pixels (selected at random) replaced with those of the training set average image. Original image attribution: Eva Rinaldi (Own work) [CC BY-SA 2.0].

We tested the model's performance with images degraded by 25%, 50%, and 75%, by projecting each test image into the PCA+LDA space in its degraded form and measuring recognition based on the nearest centroid and the nearest neighbour. We repeated this procedure across 100 iterations, each time randomly selecting the test image for each identity. Model accuracy was calculated by averaging responses across all 100 iterations, producing a proportion of iterations in which the model correctly identified the novel test image for each face from its degraded version.

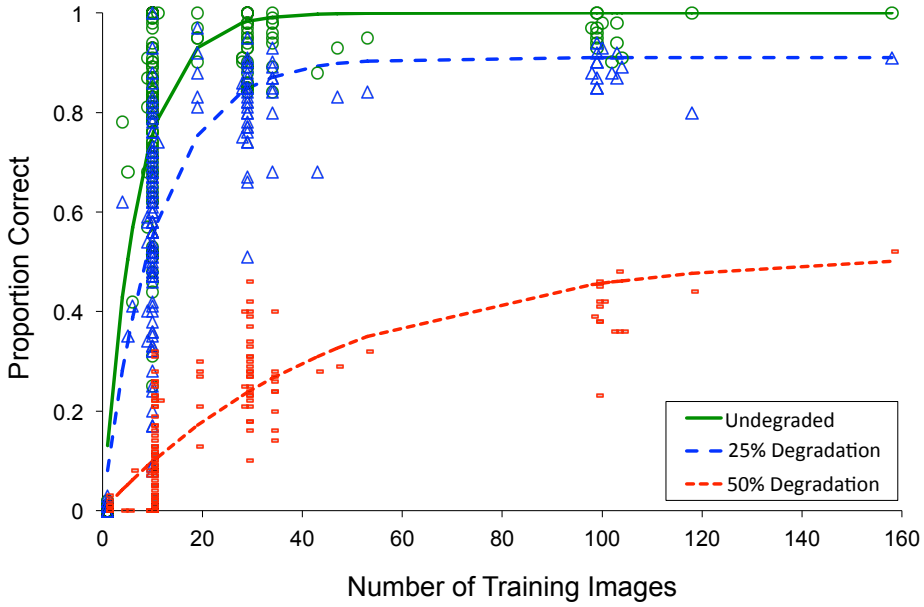

Figure 4. Effect of image degradation on recognition of untrained novel images, using the nearest centroid measure. Fitted curves represent exponential functions. Performance with the undegraded images is included for comparison. Performance following 25% and 50% degradation remains strongly influenced by the number of training images (our proxy for familiarity). At 75% degradation the model's performance is at floor, and therefore not visible.

Figure 4 shows the model's performance. To create a measure of variability in performance, data are separated into frequency bins reflecting increasing numbers of training images (1, 10, 11-20, 21-31, 31-40, 41-50 98-100, 101-110, plus two single identities with 115 and 158 images). With 25% image degradation, we found a significant relationship between face recognition accuracy and familiarity (nearest centroid, r_s = .83; nearest neighbour, r_s = .77), with well-preserved performance on the more familiar faces (i.e. those with the largest number of training images). With 50% degradation, there was a clear overall detriment, but still a significant relationship between face recognition accuracy and familiarity (nearest centroid, r_s = .83; nearest neighbour, r_s = .72). At 75%

degradation, in which the large majority of pixels carry no identity information, performance was at floor.

3.3. Recognition from internal and external features

Increasing familiarity with a face differentially enhances recognition based on its internal features such as eyes, nose and mouth, compared with external features such as hair and face shape. Although well-replicated in behavioural studies (Clutterbuck & Johnston, 2002; Ellis et al., 1979; Young et al., 1985), we are not aware of any previous attempts to simulate this pattern of increasing reliance on internal features of familiar faces.

To operationalise comparison of internal and external features in our model, we constructed images preserving only these aspects, as illustrated in Figure 5. A template for the internal feature region of the face was defined using 16 fiducial points to create a boundary around the largest area that included the eyes, nose, and mouth, whilst remaining within the overall envelope of the face outline. As all the images had been shape-normalised, the same template could always be used. This template was then used to create images in which either the external or the internal parts were replaced by the average RGB pixel values of the entire set. This manipulation results in images with the same dimensions as the whole set, which can therefore be used to test recognition in exactly the same way. However, when the internal features are replaced by the average values they become completely uninformative for face identity, and when the external features are replaced with average values then they become uninformative.

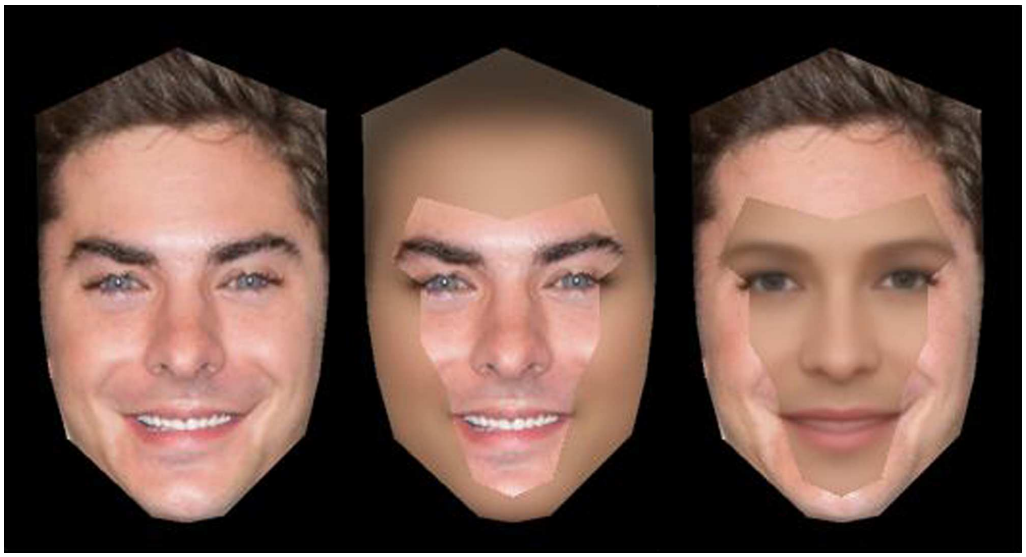

Figure 5. An illustration of the internal and external feature manipulation. A shape-standardised full-face image (left), its internal features (middle), and its external features (right). To create images with the same dimensions as the original shape-standardised image, missing regions are completed with uninformative RGB values using the average of the entire set of images. Original image attribution: Liam Mendes (Own work) [CC BY-SA 2.0].

Next, we projected each test image into the PCA+LDA space in its internal features or external features form and measured recognition. We repeated this procedure across 100 iterations, each time randomly selecting the test image for each identity. Model accuracy was calculated by averaging responses across all 100 iterations, producing a proportion of iterations in which the model correctly identified the novel image for each face based on its internal or external features. These proportions were then correlated with the familiarity of the identities (i.e., the number of images of each identity that went into the training set).

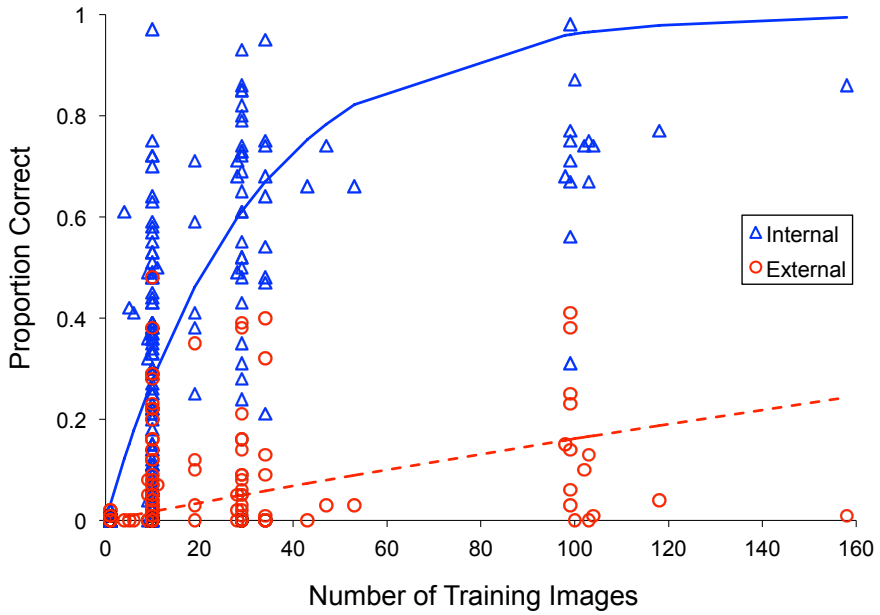

Figure 6. The proportion of correct recognitions of untrained novel face images from internal and external features using the nearest centroid measure. Each point represents the average across test images for a specific identity. Familiarity (represented by the number of training images) has more effect on recognition from internal than external features. Exponential trendlines are displayed for performance based on the internal features (solid) and the external features (dashed).

We found significant relationships between face recognition accuracy and familiarity for both internal and external features (Figure 6). However, recognition of internal features showed a stronger association with familiarity (nearest centroid, r_s = .76; nearest neighbour, r_s = .77) in comparison with recognition of external features (nearest centroid, r_s = .30; nearest neighbour, r_s = .59). So, while increased familiarity supports better recognition in general, this effect is more pronounced for the internal features, and seems to account for more of the effect of familiarity on full (unedited) face recognition shown in Figure 2.

Interestingly, these results are not simply due to the amount of pictorial information available from internal versus external features. In fact, in our cropped images, the internal features occupy 11,952 pixels, whereas the external features occupy 18,537 pixels. If all other things were equal, then, internal features could provide less RGB information regarding the individual faces than could external features, yet it was the internal features that proved most recognisable, especially for the more familiar faces. This implies both that the internal features are themselves more informative and that the PCA+LDA space has become somewhat tuned towards the use of internal features, beyond the raw amount of information available.

3.4. Face matching

3.4.1 Unfamiliar face matching

So far, we have shown that our basic model, derived from a substantial training set of highly varied naturalistic ambient images, demonstrates a graded familiarity effect in its ability to correctly 'recognise' (i.e. classify) untrained novel exemplars, and that this applies particularly to classification based on internal facial features. To test the model's applicability further, we sought to determine whether it could also fit known findings from perceptual matching tasks, where familiarity is known to exert a strong influence, with relatively poor performance in unfamiliar face matching and excellent performance with familiar faces (Bruce et al., 1999, 2001; Burton et al., 1999; Megreya & Burton, 2006, 2008).

Before turning to the role of familiarity in face matching, though, we sought first to check that the model was able to simulate performance for unfamiliar face matching. To achieve this, we used stimuli from a widely adopted standard human test of unfamiliar face matching; the Glasgow Face Matching Test (GFMT; Burton, White, & McNeill, 2010). Ability to simulate this test of unfamiliar face matching forms a starting point from which any enhancement of performance with more familiar faces can be evaluated.

The full (long) version of the GFMT (Burton et al., 2010) comprises 168 pairs of

faces, half of which match and half of which do not. Examples of images used to create the GFMT are shown in Figure 7. Participants simply indicate whether the face identities match or mismatch on each trial. The difficulty of the task stems from the fact that many image properties are unconstrained, making it hard for participants to know which image differences are relevant and which are irrelevant to the unfamiliar face identities. In the standardised version of the GFMT the images are presented in greyscale, but because of the way we implemented the current PCA+LDA model we used the original colour images for our simulation here. While using colour vs. greyscale images will undoubtedly have some effect, the role of colour in human perception of identity is limited (Kemp, Pike, White & Musselman, 1996).

Figure 7. Two example pairs of images used to create trials in the Glasgow Face Matching Test (Burton et al., 2010). The top row shows two images of different identities, while the bottom row illustrates a 'same identity' image pair. Note that all faces in the GFMT are unfamiliar and that all test items involve pairs of photographs with substantial superficial differences.

To evaluate the model's performance with the GFMT images, we created a

PCA+LDA space through training all 4,154 images in our stimulus set and then projected the pairs of images from each GFMT trial into this PCA+LDA space. Note that all faces from the GFMT are unfamiliar here – none have been used in the training set.

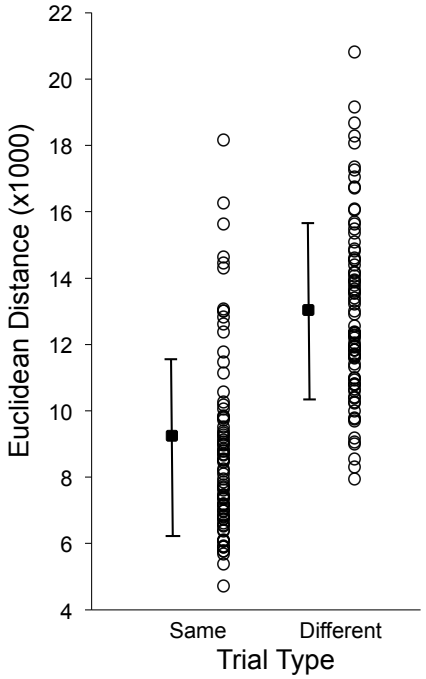

Figure 8. Performance for same identity trials and different identity trials from the Glasgow Face Matching Test (Burton et al., 2010). Each data point represents the Euclidean distance in PCA+LDA space between the pair of images forming each trial. Summary statistics (M and SD) are also displayed for each trial type.

The simplest way to represent the model's performance is in terms of the Euclidean separation between pairs of images in PCA+LDA space. This is shown in Figure 8, where the between-image distances are lower on average for same identity than for different identity pairs of unfamiliar face images, $t(166) = 10.05$, $p < .001$, Cohen's $d = 1.55$. So the model is capable of separating 'same' from 'different' identity pairs to some degree, though there is clear overlap in the distributions.

The overlapping distributions mimic what is seen in human performance on the GFMT, which is often far from perfect. With a behavioural measure of this type, in which every pair of images is physically different to some degree, human participants have to adopt their own criterion for how different the images of each face must be in order to assign them to 'same' or 'different' response categories. As Figure 8 shows, overall performance will vary according to how this criterion is set, and of course one of the key purposes of the GFMT is to measure individual differences that will in part reflect this criterion setting. In this respect, we note that calculating the distance between pairs of GFMT images and setting the 'match decision' threshold value to give comparable levels of performance across match and mismatch trials (as is observed on average with human viewers) produced performance levels of 82% and 77% accuracy for 'same' and 'different' face pairs by the model, compared to mean human performance of 92% and 88%, respectively (Burton et al., 2010). Although our computer model was trained on ambient images of many international celebrities, and has never been exposed to images of the type shown in Figure 7, it can achieve levels of matching performance within the range of human participants on these images.

As human participants show substantial individual differences on the GFMT, however, a better way to evaluate the model's performance may be in terms of whether it tends to make mistakes on the same item pairs that human observers find difficult. We found significant correlations in the expected directions involving the model distance between the image pairs and overall human performance for same trials, $r_s(82) = -0.23$, $p = .039$, and for different trials, $r_s(82) = 0.28$, $p = .009$.

3.4.2 Face matching as a function of familiarity

Having established how to apply our model to unfamiliar face matching, we investigated how an increase in familiarity affects face matching – that is, correctly perceiving that two novel instances of a face are the same person. It is well-established that face matching is easier for familiar than unfamiliar faces (Bruce et al, 2001; Johnston & Edmonds, 2009; Megreya & Burton, 2006) and so this is a key requirement for a model of familiarity. Moreover, in addition to the basic familiar/unfamiliar dichotomy,

there is already evidence that matching performance is predicted by degree of familiarity (Clutterbuck & Johnston, 2002, 2004, 2005). This makes matching a good candidate for simulation in our model.

To simulate unfamiliar face matching, we had measured the distances between pairs of images of faces that had not been explicitly represented in our model's PCA+LDA space. This is in line with the idea that *unfamiliar* face matching will rely heavily on image similarity (Hancock et al, 2000; Megreya & Burton, 2006). In contrast, *familiar* face matching need not rely much on image similarity – if two images are both recognised as Jennifer Lawrence, then they can be matched easily, regardless of their image similarity, on the basis of a more conceptual match. To simulate the impact of this conceptual matching, we investigated how the model dealt with pairs of novel images of a trained identity at different levels of familiarity.

In this simulation, we manipulated the level of familiarity of a specific face in the context of the larger model, with all its complexity and variability. To do this, we constructed variants of the model which differed only in terms of the number of items 'known' (i.e. the number of training images) for one particular individual – all the training images for the remaining 334 people remained the same in each model variant. We took the identity for whom we had the largest number of images (159 for Jennifer Lawrence) in our initial set and then varied the number of images of the actress included in the model's training set (from 0 to 151), measuring how this affected matching performance across pairs of novel images of Jennifer Lawrence's face.

For a single model iteration, we chose a random set of 151 training images plus two test images, from the 159 available for this identity. We constructed models containing the 334 other identities and incremental steps of ten training set images of Jennifer Lawrence by using 0 images or 1 image to create a baseline, and then images 1-11, 1-21, 1-31 etc. To test how familiarity (as represented by increasing the number of training images) affected face matching performance, we projected two novel test images of Jennifer Lawrence into the PCA+LDA space derived for each incremental model. For

these two novel images, we then calculated three principal measures: 1) the distance from each image to Lawrence's centroid (the distance to JL centroid measure), 2) the distance from each image to the nearest non-Lawrence centroid (the distance to nearest non-JL centroid measure), and 3) the distance between the pair of test images themselves (distance between novel images measure).

This process was repeated for 20 iterations, each time randomly selecting which images of Lawrence to use as training and novel images. We averaged across iterations and present the data in Figure 9. A comparable procedure using the nearest neighbour measure produced the same pattern.

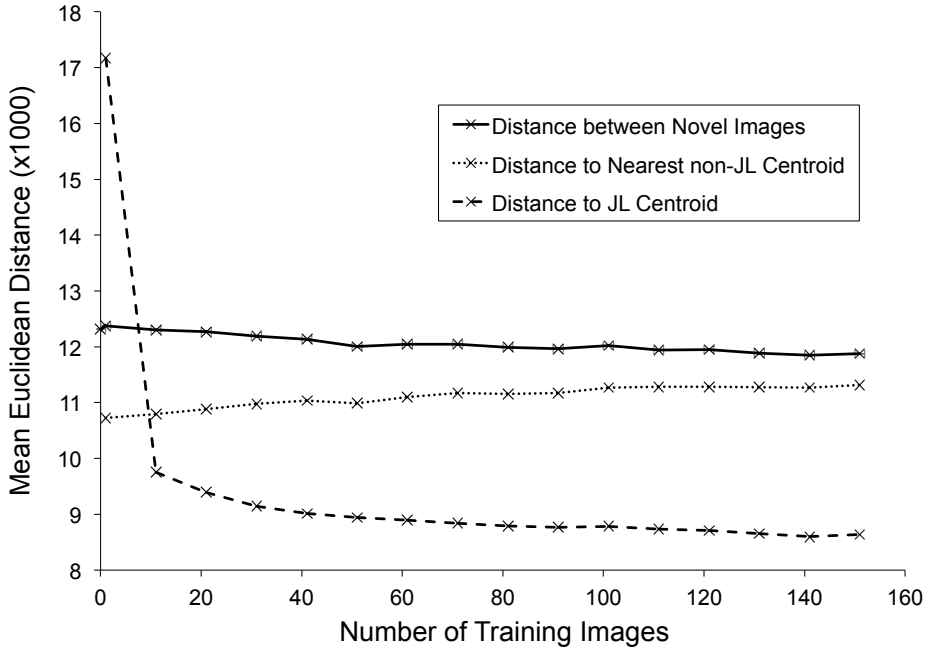

Figure 9. Performance with two novel images of Jennifer Lawrence's face as familiarity with Jennifer Lawrence (represented by the number of training images) increases. The mean Euclidean distance in PCA+LDA space (across 20 iterations) between the novel images and Lawrence's centroid, between the novel images and the nearest non-Lawrence centroid, and the distance between the two novel images themselves.

Figure 9 makes clear that the possibility of conceptual matching increases as familiarity increases, as evidenced by the increasing separation between the distance to JL centroid and distance to nearest non-JL centroid measures. Interestingly, it also shows that the centroids for faces other than JL's become more distant from the JL centroid as familiarity increases, reflecting the reshaping of the overall space produced by LDA. The Spearman correlation between familiarity (number of training images) and distance to the nearest non-JL centroid (averaged over the 20 iterations) is $r_s(14) = .95$, $p < .001$.

Less obviously, the distance between the pairs of novel images of Jennifer Lawrence also reduces slightly as familiarity increases (i.e. there is a small downward slope to the 'distance between novel images'). The Spearman correlation between familiarity (number of training images) and distance between the two novel images (averaged over the 20 iterations) is $r_s(15) = -.96$, $p < .001$. This seems to reflect a more local reshaping of the region of PCA+LDA space that represents JL's face as familiarity with her increases. This observation led us to look further at the underlying mechanisms and the extent to which they might operate in an identity-specific manner.

3.5. Underlying mechanisms

Our simulations have shown a clear advantage for increasingly familiar faces when tested using face recognition and face matching. This is consistent with key findings across decades of face research and offers insights into the nature of face familiarity. In this final empirical section, we examine underlying mechanisms in more detail.

In a previous study, we established that the combined use of LDA with PCA offers much better performance for recognising novel images of familiar identities than a PCA-based system alone (Kramer et al., 2017a). We suggested that a combination of LDA with PCA is particularly useful because each face has its own idiosyncratic forms of variability that need to be learnt as it becomes familiar (cf. Burton et al., 2016). This idiosyncratic variability limits the usefulness of generic methods such as PCA that represent only the variability of the entire set of training images without taking the idiosyncrasies of particular faces into account. Here, we put this suggestion to a formal test.

As already noted, Figure 9 shows that the between-pair distances of novel images of Jennifer Lawrence decreases slightly as the number of training images increases. That is, as the model becomes more familiar with Jennifer Lawrence's face, novel instances of her become closer together in PCA+LDA space. Although a far more gradual process than the familiarity benefits seen with face recognition (where we see a steep increase in

recognition accuracy as familiarity increases), this clustering of novel images illustrates how the underlying space in the model changes across familiarity. It is noteworthy that we should observe such a clear relationship, because the relative contribution of the Jennifer Lawrence images to the whole model is very small as training images are added; all the other 334 known people and 3,995 images remain unchanged as novel images of Jennifer Lawrence are added. The fact that the region of the PCA+LDA space that represents Jennifer Lawrence's face should change in this way underscores the key point that the variability in images of her face must be to some extent idiosyncratic and therefore needs to be represented separately from the other 3,995 images in the set (Burton et al., 2016).

To demonstrate more formally how the combination of PCA with LDA reshapes the space corresponding to each familiar identity, we used the face of Ryan Reynolds, for whom there are 104 images in our initial database. We removed 4 randomly selected images and then randomly split the remaining images into sets of 80 training images and 20 test images. The 80 training images of Ryan Reynolds were then included alongside all of the remaining 4,050 images of all other identities at the PCA stage. At the subsequent LDA stage, however, the 80 images of Ryan Reynolds were either left in (the trained identity condition) or left out (the untrained identity condition). We then calculated all pairwise distances between the 20 novel images of Reynolds after projecting these into the PCA+LDA space, and compared the mean of those distances across the trained identity and untrained identity conditions. This procedure was repeated across 20 iterations involving different random samples of 80 and 20 images from the Ryan Reynolds set.

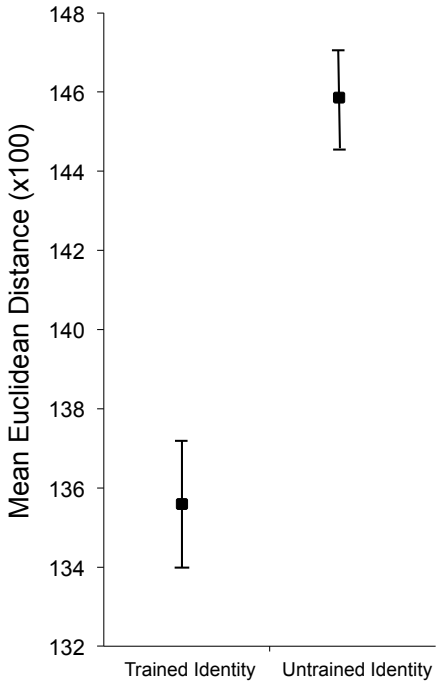

Figure 10. The mean pairwise Euclidean distances between 20 novel images of Ryan Reynolds, averaged across 20 iterations. In the trained identity condition, 80 training images of Reynolds were included in the full PCA+LDA process. In contrast, in the untrained identity condition, the 80 training images were included at the PCA stage but then removed before LDA was carried out. Error bars depict the standard error of the mean.

Results are presented in Figure 10, which shows that the distance between novel images of Ryan Reynolds is reduced by training with other images of him using the PCA+LDA procedure (the trained identity condition), in comparison to when images of Reynolds were included only in the PCA stage (the untrained identity condition). Note that the data in Figure 10 involve entirely novel images that were not used in the PCA or LDA stages, but that exactly the same sets of novel images are tested across the trained identity and untrained identity conditions. The difference between these conditions is therefore entirely attributable to the consequences of training (or not training) the identity using a different set of images at the LDA stage. Clearly, then, identity training with

LDA has the effect of reshaping the underlying PCA-based space in a way that brings any instances of the trained face into closer proximity to each other.

The benefit accruing to familiar face recognition from LDA, as an example clustering algorithm, is therefore clearly established relative to the unsupervised statistical analysis offered by PCA alone. However, our approach also allows us to go further and ask whether the representations of previously unseen faces derive *any* benefit from being projected into a space based on familiarity with known faces. The issue is important because many researchers claim that we are generic experts at perceiving face identity (Carey, 1992), whereas the view put forward here is that this characterisation in terms of face identity expertise is correct only for familiar faces (cf. Young & Burton, 2017). In effect, we have been demonstrating through our simulations based on image statistics that someone can become a Jennifer Lawrence face expert or a Ryan Reynolds face expert without actually testing whether these forms of expertise might also to some extent enhance the perception and recognition of other, unfamiliar faces.

We therefore sought to address the issue of whether the reshaping of PCA space that results from applying LDA is *entirely* person-specific? If the reshaping is completely idiosyncratic to each known face, there will be no accrued benefit from learning several different face identities through LDA on ability to represent the identities of unfamiliar faces in PCA+LDA space, whereas if the reshaping is only partly idiosyncratic then we might expect some improvement in the ability to represent unfamiliar face identities as more familiar faces are known.

To test whether learning familiar face identities can enhance representations of unfamiliar face identities, we examined the similarity between pairs of images of the same unfamiliar face when measured within a purely PCA-based space, or when measured within a PCA+LDA space built to recognise other faces. Do we gain something from tackling unfamiliar face matching within a reshaped space derived through optimising the recognition of familiar identities?

To answer this question, we collected two images for each of 40 new identities using Google Image search, following the same guidelines described earlier. Half of these celebrities were women, and all were White. None of these identities appeared in our training set. In order to determine whether our identity-derived space produced benefits for unfamiliar face matching, we projected the 80 new images into our model's PCA+LDA-trained space. For comparison, we also projected these images into the space derived from carrying out only the PCA stage of our model. In both cases, the training set was identical, but for the PCA-alone model, identity information was not used in order to derive dimensions that best discriminated between face images.

Using the 80 novel images, we simulated 40 'same' trials with the two images of each identity. As a measure of the model's performance, we calculated the Euclidean distances in PCA space and in PCA+LDA space between these pairs of images. In order to generate 40 'different' trials, we paired one image for each identity with a foil chosen from the other images in this set. We took care to match the two faces on basic descriptors like sex, hair colour, the presence of stubble, and age, acknowledging the limitations inherent in such a small sample of faces. For each of these trials, we again calculated the Euclidean distances in PCA space and in PCA+LDA space between the two images.

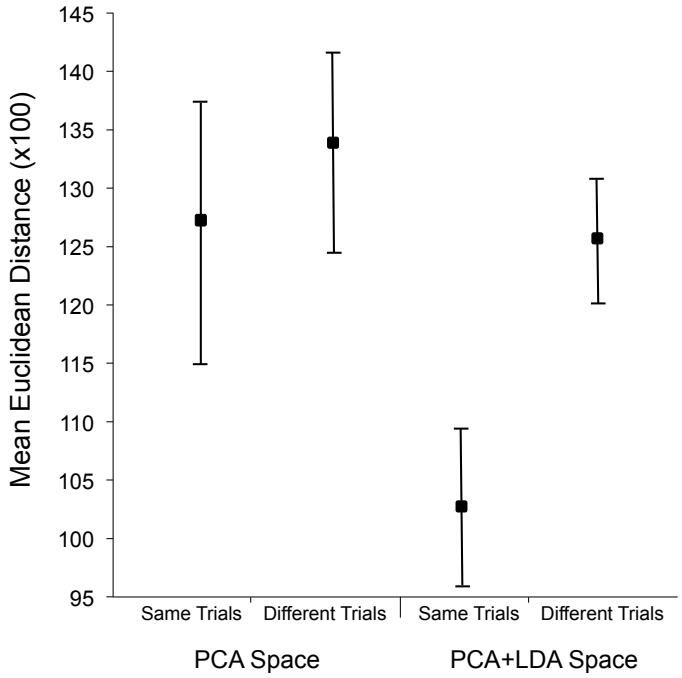

Figure 11. Mean Euclidean distances between 'same' and 'different' identity pairs of novel images of unfamiliar faces in PCA-space (image similarity only) and PCA+LDA-spaces (image similarity plus identity training for familiar faces). Error bars show 95% confidence intervals. There is a clear benefit to determining that different images of unfamiliar faces represent the same identity in the PCA+LDA space, despite it having been trained on other faces.

Figure 11 illustrates these 'same' and 'different' identity distances when the images are projected into PCA-based and PCA+LDA spaces. We find that there is no statistically significant difference between the two types of trial for the PCA-alone model, $t(78) = 1.02$, $p = .313$, Cohen's $d = 0.23$. In contrast, the PCA+LDA model successfully discriminates between 'same' and 'different' identity image pairs, $t(78) = 4.90$, $p < .001$, Cohen's $d = 1.10$.

This result is an important one. Using the same training image set, we can derive dimensions that best capture the variance in the images' pixel values (PCA) or we can

calculate dimensions that are optimal for identity discrimination (PCA+LDA). This latter case appears to result in a reshaped space well suited to discriminating the identities of novel images of both familiar (trained) faces *and*, to some extent, of unfamiliar (untrained) faces. Even though the faces used here never appeared in the training set, we find that the PCA+LDA space provides some support for establishing similarity in identity beyond superficial image similarity that extends to completely new faces.

4. Discussion

We set out to test the idea that familiarity can be thought of as involving bottom-up low level image descriptions, together with a top-down mechanism for cohering superficially variable images of the same person. Using a combination of PCA and LDA, we successfully reproduced key findings in the literature regarding how people perform in face recognition and face matching tasks. Crucially, we were able to show familiarity advantages for entirely untrained images with a model based only on the optimal separation of identities in PCA+LDA space. The benefits of increasing familiarity - as defined by the number of different training images used for a given face - accrued in terms of better recognition of novel exemplars of the trained faces, better face matching, better resistance to the effects of image degradation, and better recognition from internal than external features.

The combination of bottom-up image description, with top-down clustering, has been used in previous models (e.g. Bekios-Calfa et al, 2011). In our own previous work (Kramer et al, 2017a) we have used this approach to classify sex and race from familiar and unfamiliar faces. However, previous models, including our own, treated familiarity as a bivalent variable, in which 'familiar' faces were uniformly familiar. For the reasons described above, that is an incomplete approach to understanding familiarity which is self-evidently graded, i.e. we know some faces better than others. Here we show that the same general approach used in understanding other aspects of face perception, can also be used to begin to understand the much more complex nature of familiarity itself.

The contrast between familiar and unfamiliar faces is often linked, either explicitly or by implication, to the idea of qualitative differences between the processing of familiar and unfamiliar faces (Bruce & Young, 1986; Burton, Bruce & Hancock, 1999; Hancock et al., 2000; Megreya & Burton, 2006). For example, the perception of faces that have never been encountered before is so image-dependent that participants experience little difficulty in being taught the incorrect information that two different images of the same face belong to different people (Longmore et al., 2008). Yet at the other extreme, the recognition of highly familiar faces is so fluent that we can find it hard to see how different two images of a familiar person actually are (Jenkins et al., 2011) and relatively difficult to remember the details of specific images that have been seen (Armann, Jenkins & Burton, 2016). Our results show how despite familiarity lying along a graded continuum it remains reasonable to look upon the extremes of familiarity as involving differences that are to all intents and purposes so large as to appear qualitative in nature, but they also show how there can be gradations in performance between these extremes.

Given the highly unconstrained nature of the images used in the simulations above, sampled from internet search and with no control of low-level image properties, the performance we report is surprisingly good, as well as having human-like properties. Of course, we do not wish to claim that the human brain explicitly uses PCA or LDA. Instead, the model presented here demonstrates that a clustering algorithm, cohering together multiple instances of the same person, can use simple intensity (pixel) level statistical structure to deliver apparently high-level information in the form of face recognition. The model provides an existence proof of this, without commitment to specific implementation.

None the less, one might ask how far learning faces from photographs is truly representative of natural face learning? At the present state of knowledge we cannot be completely certain, but some points stand out. First, although photographs are entirely static, recognition of static images is so good that any idiosyncratic patterns of facial movement convey no measurable benefit under normal circumstances; substantial image

degradation is needed before any effects of facial movement become apparent (Lander, Christie & Bruce, 1999; O'Toole, Roark & Abdi, 2002). Second, while faces learnt from single photographs show remarkably poor generalisation (Bruce, 1982; Longmore et al., 2008), it is none the less clear that faces can be learnt from multiple variable photographs of the same face in ways that show properties comparable to natural recognition (S. Andrews et al., 2015; Dowsett et al., 2016). Moreover, the degree of variability in exposure to such multiple images is predictive of how well this learning can generalise to new exemplars (Ritchie & Burton, 2017). Both phenomena fit with the observation that variability in the views of faces to which we are exposed is typical of our everyday lives. On balance, then, the available evidence suggests that there is nothing special (or unrepresentative) about learning faces from photos.

In common with other graphical approaches, we began by standardising the positions of key fiducial positions in each image (Beymer, 1995 Craw, 1995; Vetter & Troje, 1995). Behavioural studies show that such stimuli remain easily recognisable to human observers (Burton et al., 2005; T. Andrews et al., 2016). The analyses were then conducted entirely on pixel-based surface colour and brightness values. These surface properties involve a combination of the surface reflectances of different parts of the face (known as its albedo map in the computer science literature), prevailing illumination conditions (such as direction of lighting) and camera characteristics. In any given image, there will be an unspecified mix of these different factors. Importantly, our model did not make any attempt explicitly to represent shape information concerning the second-order configuration of features, three-dimensional information about head shape, knowledge of how expressions can alter the face, and other visual properties often thought to be involved in face recognition. Indeed, with these highly variable everyday images we found that 2D shape information from the fiducial locations alone was of limited value in comparison to the surface properties of the shape-normalised images. This does not mean that shape is irrelevant; there is evidence that human observers can be sensitive to shape properties (O'Toole, Vetter & Blanz, 1999; T. Andrews et al., 2016). In this respect we note that some shape information will still be available in the standardised images via patterns of shape from shading, texture changes due to opening or closing the mouth and

eyes, and so on (cf. Sormaz, Young & Andrews, 2016). Whether or not this is the underlying cause, the simulations show that learning from covariation of surface information within and between identities is sufficient to underpin human-like performance.

Our simulation exploring the importance of internal versus external facial features proved consistent with the behavioural evidence that people rely to a greater extent on the internal features for more familiar faces (Clutterbuck & Johnston, 2002; Ellis et al., 1979; Young et al., 1985). While we found that in general novel images of more familiar identities were better recognised when projected into the PCA+LDA space, this relationship between familiarity and accurate recognition was stronger for internal features than external features.

This ability to simulate the importance of the internal features for familiar face recognition directly addresses an important debate concerning the origins of this finding, which has been interpreted in two very different ways. The interpretation originally offered by Ellis et al. (1979) was that the internal facial features receive most attention in social encounters because of their critical role in social signals such as gaze and facial expression. They therefore become differentially represented for familiar faces because these are the features that have been most looked at (Ellis et al., 1979). In contrast, an alternative interpretation offered by Young et al. (1985) was that while external features, and especially the hair, can often dominate any particular photo of an unfamiliar person, this is not a very diagnostic feature of identity, because it is easily changeable. Therefore, over increasing exposure, people may come to rely on aspects of the face which change less across encounters (Young, 1984; Young et al., 1985; Bonner, Burton, & Bruce, 2003; Osborne & Stevenage, 2008). In fact, by presenting only the internal features, researchers have been able to improve unfamiliar face learning (Longmore, Liu, & Young, 2015) and matching accuracy in some conditions (Kemp, Caon, Howard, & Brooks, 2016).

These interpretations of the internal feature advantage for familiar faces differ in the emphasis they place on properties that are intrinsic to how images of faces themselves vary in the everyday world (the 'image-based' interpretation favoured by Young et al., 1985) or on the way these facial images are analysed by human perceivers (the more 'social' interpretation suggested by Ellis et al., 1979). Our data were consistent with the key prediction of the image-based account, that differential salience of the internal features will accrue to familiar faces simply on the basis of the nature of everyday image variability. Young et al. (1985) had in fact suggested that it might be possible to tease apart image-based and more social explanations "by studying how the differential salience of the internal features is established as faces become increasingly familiar" (Young et al., 1985, p.745). Whilst PCA+LDA is clearly not intended as a full model of brain processes involved in face recognition, it does offer an effective way of finding the information sufficient to support recognition.

From a more general perspective, our simulations show how increasing familiarity with a face leads to better performance. This needs to be considered with respect to previous suggestions that averaged images can capture the essential invariant characteristics of a specific face identity by eliminating identity-irrelevant variability between images (Burton et al., 2005; Jenkins & Burton, 2008). We do not wish to deny the value of that observation, but it is important to appreciate that this is *not* how the present approach works. Instead, rather than seeking to average away image variability, what we do here is to make use of it. What LDA achieves is to maximise between-identity distances (the separation between images of different faces) while minimising the within-identity distances (by clustering images of the same face close together). Faces that include more images in the training set will therefore have a greater influence on the resulting dimensions, but averages are never calculated by the model (though our centroid measure of its performance involves an averaged location in the representational space).

We think that LDA may be particularly useful in this respect because each face has its own idiosyncratic forms of variability across different image views (cf. Burton et al.,

2016). As we noted, whereas many researchers claim that we are generic experts at perceiving face identity (Carey, 1992), the view put forward here is that this characterisation is mainly correct for familiar faces (Young & Burton, 2017). Figures 9 and 10 show how training the identity of a familiar face with LDA reshapes the underlying PCA-based space in a way that brings entirely novel instances of the trained face into closer proximity to each other than they would be from their image descriptions alone. This observation emphasises the importance of supervised learning to finding identity-specific variability. An approach based purely on an unsupervised analysis of the image statistics of the perceptual input alone (i.e. PCA of the image training set) does not do so well (Figure 11). This is consistent with studies of human face learning, in which expectations about identity (e.g. how many individuals to expect in a set of faces) has a marked influence on the perception of identity (Andrews et al, 2015; Menon, White & Kemp, 2015b).

Taking the question of underlying mechanisms a step further, however, we were also able to demonstrate that LDA reshapes the underlying PCA-based space in a way that confers some benefit to representing the identities of entirely unfamiliar (untrained) faces (see Figure 11). This result speaks to a long-standing problem in face research – the extent to which general processes operate when recognising faces. At one extreme, images of unfamiliar faces have been held to be unable to recruit privileged or special processing available to familiar faces (Megreya & Burton, 2006; see also Hancock et al., 2000). From the simulations presented here, this now seems too strong a claim, though it remains the case that models of face processing which ignore pervasive differences between familiar and unfamiliar faces are inadequate (Young & Burton, 2017). In the PCA+LDA approach, we seem to have a useful integration. Familiar faces shape similarity space in a way which benefits them optimally, but which also provides some benefit to unfamiliar face processing. Hence although our primary expertise in face recognition is for the identities of familiar faces, this has consequences for recognition of unfamiliar faces too.

The fact that LDA reshapes the underlying PCA-based space has profound

implications for the widely-used face space metaphor. Face space is conceived as a set of hypothetical multidimensional linear vectors that represent the differences between faces (Valentine, 1991, 2001). Face space models then try to represent each face identity as a discrete point in this multidimensional space, noting that some faces will be closer together or further apart from each other. Although the dimensions of face space remain unspecified, the underlying assumption of linearity has strong parallels with PCA approaches and has been shown to approximate cell responses in the monkey brain (Chang & Tsao, 2017). However, our demonstrations here show that a completely linear space based on image properties alone does not cope well with the problem of within-person variability and that LDA can be used to reshape the space into something more useful. If we use a pretentious analogy, the presence of a highly familiar face distorts space in a way that resembles a large mass distorting the space around it in Einstein's theory of relativity. A related point concerning distortions of a hypothetical face space created by familiar attractors had been suggested by Tanaka et al. (1998).

In sum, our aim here was to present a model of how face familiarity might be conceptualised. We have presented simulations that show our model performs realistically on face recognition and matching tasks, with increasingly familiar faces being better matched and recognised, showing resistance to degradation, and increasing dependence on their internal features. To our knowledge, we are the first to model varying degrees of face familiarity in a single system, and to explore how well such a system can encompass established results from a range of key findings based on data for human participants. We hope to have taken the first steps towards providing a working account of the mechanisms that exploit face variability to achieve familiarity.

Funding

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement n.323262 to AMB, and from the Economic and Social Research Council, UK [ES/J022950/1] to AMB.

References

Andrews, S., Jenkins, R., Cursiter, C., & Burton, A. M. (2015). Telling faces together: Learning new faces through exposure to multiple instances. *Quarterly Journal of Experimental Psychology, 68*, 2041–2050.

Andrews, T. J., Baseler, H., Jenkins, R., Burton, A. M., & Young, A. W. (2016). Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. *Cortex, 83*, 280-291.

Armann, R. G. M., Jenkins, R., & Burton, A. M. (2016). A familiarity disadvantage for remembering specific images of faces. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 571-580.

Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2011). Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 858–864.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711–720.

Beveridge, J. R., Phillips, P. J., Givens, G. H., Draper, B. A., Teli, M. N., & Bolme, D. S. (2011). When high-quality face images match poorly. In *Proceedings of the ninth international conference on automatic face and gesture recognition* (pp. 572–578). Santa Barbara, CA: IEEE.

Beymer, D. (1995). Vectorizing face images by interleaving shape and texture computations. MIT AI Lab memo 1537. Cambridge, MA: Massachusetts Institute of Technology.

Bonner, L., Burton, A. M., & Bruce, V. (2003). Getting to know you: How we learn new faces. *Visual Cognition, 10*(5), 527-536.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105-116.

Bruce, V. (1994). Stability from variation: The case of face recognition. *Quarterly Journal of Experimental Psychology, 47*, 5-28.

Bruce, V., & Young, A. W. (1986). Understanding face recognition. *British Journal of*

Psychology, *77*(3), 305–327.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5*(4), 339-360.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied, 7*(3), 207–218.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*(8), 1467-1485.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: a model of familiar face recognition. *Cognitive Science, 23*, 1-31.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*, 256-284.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology, 102*, 943-958.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science, 40*(1), 202-223.

Burton, A. M., Miller, P., Bruce, V., Hancock, P. J. B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image formats. *Vision Research, 41*(24), 3185–3195.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science, 10*(4), 482-496.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1), 286-291.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science, 10*(3), 243-248.

Carey, S., & Diamond, R. (1977). From piecemeal to configurational representation of

faces. *Science, 195*, 312-314.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society, London, B: Biological Sciences, 335*, 95-103.

Chang, L., & Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell, 169*, 1013-1028.

Chen, L. F., Liao, H. Y. M., Ko, M. T., Lin, J. C., & Yu, G. J. (2000). New LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, *33*(10), 1713–1726.

Clutterbuck, R., & Johnston, R. A. (2002). Exploring levels of face familiarity by using an indirect face-matching measure. *Perception, 31*, 985-994.

Clutterbuck, R., & Johnston, R. A. (2004). Matching as an index of face familiarity. *Visual Cognition, 11*(7), 857-869.

Clutterbuck, R., & Johnston, R. A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. *European Journal of Cognitive Psychology, 17*(1), 97-116.

Craw, I. (1995). A manifold model of face and object recognition. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp.183-203). London: Routledge.

Dowsett, A. J., Sandford, A., & Burton, A. M. (2016). Face learning with multiple images leads to fast acquisition of familiarity for specific individuals. *Quarterly Journal of Experimental Psychology, 69*(1), 1-10.

Eger, E., Schweinberger, S. R., Dolan, R. J., & Henson, R. N. (2005). Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. *NeuroImage, 26*, 1128-1139.

Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *Quarterly Journal of Experimental Psychology, 39A*, 193-210.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8*(4), 431–439.

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *i-Perception, 5*, 589-601.

Fisher, R.A. (1936). The use of multiple measures in taxonomic problems. *Annals of Eugenics, 7*(2), 179-188.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330-337.

Hay, D. C. (2000). Testing instance models of face repetition priming. *Memory & Cognition, 28(2)*, 192-203.

Hill, H., & Bruce, V. (1996). Effects of lighting on matching facial surfaces. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 986-1004.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*, 1221-1240.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. Technical Report 07-49, University of Massachusetts, Amherst.

Hussain, Z., Sekuler, A. B., & Bennett, P. J. (2009). Perceptual learning modifies inversion effects for faces and textures. *Vision Research*, *49*(18), 2273–84.

Itz, M. L., Golle, J., Luttmann, S., Schweinberger, S. R., & Kaufmann, J. M. (2017). Dominance of texture over shape in facial identity processing is modulated by individual abilities. *British Journal of Psychology*, *108*, 369–396.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science, 319*, 435.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B, 366*, 1671-1683.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313–323.

Jing, X.-Y., Wong, H.-S., & Zhang, D. (2006). Face recognition based on 2D Fisherface approach. *Pattern Recognition*, *39*(4), 707–710.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577-596.

Kemp, R. I., Caon, A., Howard, M., & Brooks, K. R. (2016). Improving unfamiliar face matching by masking the external facial features. *Applied Cognitive Psychology, 30*(4), 622-627.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces - the role of shape from shading and pigmentation cues. *Perception, 25*, 37-52.

Klatzky, R. L., & Forrest, F. H. (1984). Recognizing familiar and unfamiliar faces. *Memory & Cognition, 12*(1), 60–70.

Kramer, R. S. S., Young, A. W., Day, M. G., & Burton, A. M. (2017a). Robust social categorization emerges from learning the identities of very few faces. *Psychological Review, 124*(2), 115-129.

Kramer, R. S. S., Jenkins, R., & Burton, A. M. (2017b). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior Research Methods, 49,* 2002-2011.

Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory & Cognition, 27*, 974-985.

Liu, C. H., Bhuiyan, A., Ward, J., & Sui, J. (2009). Transfer between pose and illumination training in face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35(4)*, 939-947.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 77-100.

Longmore, C. A., Liu, C. H., & Young, A. W. (2015). The importance of internal facial features in learning new faces. *Quarterly Journal of Experimental Psychology, 68*(2), 249-260.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied, 14*(4), 364–372.

Menon, N., White, D., & Kemp, R. I. (2015a). Variation in photos of the same face drives improvements in identity verification. *Perception, 44*(11), 1332-1341.

Menon, N., White, D., & Kemp, R. I. (2015b). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. *Quarterly Journal of Experimental Psychology*, *68*(9), 1777-1793.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance, 41(3)*, 577-581.

O'Toole, A. J., Edelman, S., & Bülthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38*, 251-263.

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences, 6,* 261-266.

O'Toole, A. J., Vetter, T., & Blanz, V. (1999). Three-dimensional shape and two-dimensional surface reflectance contributions to face recognition: an application of three-dimensional morphing. *Vision Research, 39*, 3145-3155.

Osborne, C. D., & Stevenage, S. V. (2008). Internal feature saliency as a marker of familiarity and configural processing. *Visual Cognition, 16*(1), 23-43.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3(4)*, 406-417.

Read, J. D., Vokey, J. R., & Hammersley, M. (1990). Changing photos of faces: Effects of exposure duration and photo similarity on recognition and the accuracy--confidence relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 870-882.

Reynolds J.K., & Pezdek, K. (1992). Face recognition memory: The effects of exposure duration and encoding instruction. *Applied Cognitive Psychology, 6*, 279-292.

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology, 70,* 897-905.

Sormaz, M., Young, A. W., & Andrews, T. J. (2016). Contributions of feature shapes and surface cues to the perception of facial expressions. *Vision Research, 127*, 1-10.

Tanaka, J., Giles, M., Kremen, S., & Simon, V. (1998). Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition, 68*. 199-220.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology, 43A*, 161-

204.

Valentine, T. (2001). Face-space models for face recognition. In M. J. Wenger & J. T. Townsend (Eds.) *Computational, geometric and process perspectives on facial recognition* (pp. 83-113). Mahwah, NJ: Erlbaum.

Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*(5), 525–535.

Vetter, T., & Troje, N. (1995). Separation of texture and two-dimensional shape in images of human faces. In G. Sagerer, S. Posch, & F. Kummert (Eds.), Mustererkennung 1995,

Yarmey, A. D. (1971). Recognition memory for familiar "public" faces: Effects of orientation and delay. *Psychonomic Science, 24*, 286-288.

Young, A.W. (1984). Right cerebral hemisphere superiority for recognizing the internal and external features of famous faces. *British Journal of Psychology, 75*, 161-169.

Young, A. W., & Bruce, V. (2011). Understanding person perception. *British Journal of Psychology, 102*, 959-974.

Young, A. W., & Burton, A. M. (1999). Simulating face recognition: implications for modelling cognition. *Cognitive Neuropsychology, 16*, 1-48.

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science, 26,* 212-217.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14*, 737-746.